

## PATENT ABSTRACTS OF JAPAN

(11)Publication number : 2002-358089

(43)Date of publication of application : 13.12.2002

(51)Int.Cl.

G10L 11/02

G10L 15/04

G10L 15/28

(21)Application number : 2001-166972

(71)Applicant : DENSO CORP

(22)Date of filing : 01.08.2001

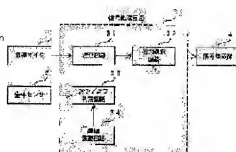
(72)Inventor : TSURUTA KAZUHIRO  
YAMADA YOSHIYASU

## (54) METHOD AND DEVICE FOR SPEECH PROCESSING

(57)Abstract:

PROBLEM TO BE SOLVED: To provide a method and a device for speech processing which properly process an inputted speech signal to facilitate recognition of the speech.

SOLUTION: An on/off discrimination circuit 33 of a signal processing circuit 3 performs discrimination with a set threshold on the basis of a signal from a living body sensor 2 as a bone conduction microphone and detects a duration section of bone conduction sounds, namely, a speaking section of a user to output a signal showing turning-on. In the section where the signal showing turning-on is outputted, a voice signal from an air conduction microphone 1 is outputted as it is by an output selection circuit 32. While a signal showing turning-off is outputted from the on/off discrimination circuit 33, namely, in a section other than the speaking section, the output selection circuit 32 reduces the gain to output the voice signal from the air conduction microphone 1.



## LEGAL STATUS

[Date of request for examination]

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

[Date of extinction of right]

(51) Int.Cl. <sup>7</sup>	識別記号	F I	キーワード <sup>*</sup> (参考)	
G 1 0 L	11/02	G 1 0 L 3/00	5 1 3	5 D 0 1 5
	15/04		5 1 1	
	15/28			

審査請求 未請求 請求項の数12 O L (全 7 頁)

(21) 出願番号 特願2001-166972(P2001-166972)

(22) 出願日 平成13年6月1日 (2001.6.1)

(71) 出願人 000004260

株式会社デンソー

愛知県刈谷市昭和町1丁目1番地

(72) 発明者 鶴田 和弘

愛知県刈谷市昭和町1丁目1番地 株式会  
社デンソー内

(72) 発明者 山田 芳靖

愛知県刈谷市昭和町1丁目1番地 株式会  
社デンソー内

(74) 代理人 100082500

弁理士 足立 敏

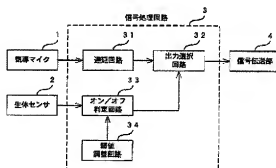
Fターム(参考) 5D015 D002 D003

## (54) 【発明の名称】 音声処理装置及び音声処理方法

## (57) 【要約】

【課題】 入力される音声信号を適切に処理し、その音声の認識を容易にするための音声処理装置及び音声処理方法を提供する。

【解決手段】 骨伝導マイクロホンである生体センサ2からの信号に基づいて、信号処理回路3のオン/オフ判定回路33が、設定された閾値での判断を行い、骨伝導音の継続区間、すなわち利用者の発声区間を検知してオンを示す信号を出力する。そして、出力選択回路32により、オンを示す信号が出力されている区間では、気導マイクロホン1からの音声信号がそのまま出力される。一方、オン/オフ判定回路33からオフを示す信号が出力されている間、つまり、発声区間に対応しない区間においては、出力選択回路32が、利得を下げて気導マイクロホン1からの音声信号を出力する。



## 【特許請求の範囲】

【請求項1】 空気振動を検知して話者の音声信号化する気導マイクロホンと、話者の発声動作を検知する生体センサと、前記生体センサからの信号に基づき、話者が実際に発声している区間である発声区間を検知する検知手段と、前記気導マイクロホンからの音声信号のうち、前記検知手段にて検知された発声区間に対応する区間の音声信号を出力する出力手段とを備えることを特徴とする音声処理装置。

【請求項2】 請求項1に記載の音声処理装置において、前記出力手段は、前記発声区間に対応しない区間の前記気導マイクロホンからの音声信号については利得を下げて出力することを特徴とする音声処理装置。

【請求項3】 請求項1又は2に記載の音声処理装置において、前記検知手段は、前記生体センサからの信号値が所定の閾値を上回ったことを判断して前記発声区間を検知することを特徴とする音声処理装置。

【請求項4】 請求項3に記載の音声処理装置において、前記検知手段は、前記生体センサからの信号値が所定の閾値を上回っている区間を前記発声区間とすることを特徴とする音声処理装置。

【請求項5】 請求項3に記載の音声処理装置において、前記検知手段は、前記生体センサからの信号値が所定の閾値を上回ったことを判断してから一定区間を前記発声区間とすることを特徴とする音声処理装置。

【請求項6】 請求項1～5のいずれかに記載の音声処理装置において、

前記生体センサは、骨伝導マイクロホン、超音波センサ、赤外線センサ、筋電位センサ、加速度センサ、又は、光式振動ピックアップの少なくともいずれか一つを用いて構成されていることを特徴とする音声処理装置。

【請求項7】 空気振動を検知する気導マイクロホンを通じて音声信号を入力すると共に、話者の発声動作を検知する生体センサからの信号を入力し、

前記生体センサからの信号に基づき話者の発声区間を特定し、

前記気導マイクロホンからの音声信号のうち、当該特定された発声区間に対応する区間の音声信号を出力することを特徴とする音声処理方法。

【請求項8】 請求項7に記載の音声処理方法において、前記特定された発声区間に対応しない区間の気導マイクロホンからの音声信号については利得を下げて出力することを特徴とする音声処理方法。

【請求項9】 請求項7又は8に記載の音声処理方法において、前記生体センサからの信号値が所定の閾値を上回ったことを判断して前記発声区間を特定することを特徴とする音声処理方法。

【請求項10】 請求項9に記載の音声処理方法において、前記生体センサからの信号値が所定の閾値を上回っている区間を前記発声区間とすることを特徴とする音声処理方法。

【請求項11】 請求項9に記載の音声処理方法において、前記生体センサからの信号値が所定の閾値を上回ったことを判断してから一定区間を前記発声区間とすることを特徴とする音声処理方法。

【請求項12】 請求項7～11のいずれかに記載の音声処理方法において、前記生体センサは、骨伝導マイクロホン、超音波センサ、赤外線センサ、筋電位センサ、加速度センサ、又は、光式振動ピックアップの少なくともいずれか一つを用いて構成されていることを特徴とする音声処理方法。

## 【発明の詳細な説明】

## 【0001】

【発明の属する技術分野】 本発明は、音声の認識を容易にするための音声処理技術に関する。

## 【0002】

【従来の技術】 音声認識技術においては、認識対象とする音声区間を適切に特定することが重要である。それは、特定された音声区間の音声信号を認識対象語と比較するなどして音声認識がなされるからであり、音声区間に余分な音声や雑音が含まれると、認識率の著しい低下を招くためである。つまり、認識対象となる音声区間は、話者が実際に声を出している区間（以下「発声区間」という。）に等しくなっていることが望ましい。

【0003】 従来、音声区間の特定は、トークスイッチと呼ばれるスイッチのオン・オフによる利用者からの指示でなされたり、また、入力された音声信号のレベルに基づいてなされたりしていた。

## 【0004】

【発明が解決しようとする課題】 しかしながら、トークスイッチを使用する場合、トークスイッチのオンとなっている期間（音声区間）が必ずしも、発声区間であるとは限らない。例えば、トークスイッチを押下した後、一呼吸おいてから話者が発声することもあり得るし、特に、複数の連続した単語を順に認識するような場合、ある単語と別の単語との切れ目が存在するからである。このとき、話者の音声と共に外界音をマイクロホンが拾ってしまうと、特定された音声区間を対象とする音声認識において、その精度が極端に低下してしまう。

【0005】 一方、音声信号のレベルに基づいて発声区間を特定する場合でも同様であり、外界音のレベルが高くなると発声区間以外を音声区間と判定してしまうことがある。なお、ここでは音声認識の例に挙げて説明したが、外界音が比較的大きな場所において、電話をした

る。すなわち、発声区間以外の部分の外界音によって、話者の音声が大変聞き取り難いものになる。

【0006】本発明は、上述した問題点を解決するためになされたものであり、入力される音声信号を適切に処理し、その音声の認識を容易にするための音声処理装置及び音声処理方法を提供することを目的とする。

【0007】

【課題を解決するための手段及び発明の効果】上述した目的を達成するためになされた請求項1に記載の音声処理装置では、空気振動を検知する気導マイクロホンにて話者の音声信号が信号化されるのであるが、ここで、生体センサが話者の発声動作を検知し、検知手段が、生体センサからの信号に基づき、話者が実際に発声している区間である発声区間を検知する。そして出力手段によって、上述した気導マイクロホンからの音声信号のうち、検知された発声区間に対応する区間の音声信号が出力される。

【0008】これによって、話者が実際に発声している区間、すなわち発声区間の音声信号が本装置から出力されることになり、この音声信号を音声認識装置、電話機、あるいは、録音装置への入力信号とすれば、話者の音声の認識が容易になる。特に、連続して入力される音声信号のうちの発声区間部分のものを出力できるため、対話といった連続した単語の音声認識を行う音声認識装置への入力信号とすれば、本発明の効果はより一層顕著なものとなる。

【0009】このとき、気導マイクロホンからの信号のうち、検知された発声区間に対応する信号のみを出力するようにしてもよい。しかし、検知された区間だけで音声信号を出力すると、区間の境界部分に外界音などが存在する場合、区間の境界部分で音声信号が切断されることにより、区間の境界部分の音声信号の波形が切り立って、高周波帯域の雑音が音声信号に混入する可能性が高くなる。

【0010】そこで、請求項2に示すように、出力手段は、発声区間に対応しない区間の信号については利得を下げて出力する構成にするといよい。このようにすれば、区間の境界部分に外界音などが存在しても、雑音が混入する可能性を低減でき、さらに、音声の認識が容易になる。

【0011】ところで、検知手段が生体センサからの信号に基づいて発声区間を検知することは既に述べた。例えば、請求項3に示すように、検知手段は、生体センサからの信号値が所定の閾値を上回ったことを判断して発声区間を検知することが考えられる。さらに具体的に言えば、請求項4に示すように、生体センサからの信号値が所定の閾値を上回っている区間を発声区間とすることが考えられる。ただし、生体センサからの信号が音声信号などの疎密波であるとき、信号が零点を通るために、発声区間の途中で何度も閾値を跨ぐことになる。そこでこ

の場合は、請求項5に示すように、生体センサからの信号値が所定の閾値を上回ったことを判断してからの一定区間を発声区間とすることが考えられる。

【0012】このような生体センサは、例えば請求項6に示すように、骨伝導マイクロホン、超音波センサ、赤外線センサ、筋電位センサ、加速度センサ、又は、光式振動ピックアップのすくなくともいずれか一つを用いて構成することが考えられる。以上は、音声処理装置の発明として上述してきたが、音声処理方法の発明として実現することもできる。

【0013】すなわち、請求項7に示す音声処理方法では、空気振動を検知する気導マイクロホンを通じて音声信号を入力すると共に、話者の発声動作を検知する生体センサからの信号を入力し、生体センサからの信号に基づき話者の発声区間を検知する。そして、気導マイクロホンからの音声信号のうち、当該検知された発声区間に対応する区間の音声信号を出力する。

【0014】このような音声処理方法によれば、上述した音声処理装置と同様の効果が得られる。つまり、発声区間の音声信号が出力されるため、この音声信号に基づけば、音声の認識を容易にできる。このような音声処理方法においても、上述した音声処理装置と同様に、発声区間に対応しない区間の音声信号については利得を下げて出力するようにしよい（請求項8）。このようにすれば、区間の境界部分に外界音などが存在しても、雑音が混入する可能性を低減でき、さらに、音声の認識が容易になる。

【0015】また、生体センサからの信号値が所定の閾値を上回ったことを判断して発声区間を検知するように構成することが考えられる（請求項9）。このときは、生体センサからの信号値が所定の閾値を上回っている区間を発声区間とする構成にしたり（請求項10）、生体センサからの信号値が所定の閾値を上回ったことを判断してからの一定区間を発声区間とする構成にしたりすることが考えられる（請求項11）。

【0016】さらに生体センサを、骨伝導マイクロホンや、超音波センサ、赤外線センサ、筋電位センサ、加速度センサ、光式振動ピックアップのすくなくともいずれか一つを用いて実現することも同様である（請求項12）。なお、上述してきた音声処理装置や音声処理方法は、従来のトークスイッチなどと併用して用いられることも考えられるが、本装置や本方法を応用して自動的に音声区間を認識するトークスイッチを構成することもできる。

【0017】

【発明の実施の形態】以下、本発明を具体化した一実施例を図面を参照して説明する。図1は、実施例の信号処理装置の概略構成を示すブロック図である。本信号処理装置は、気導マイクロホン1と、生体センサ2と、信号処理回路3と、信号伝送部4とを備えている。そして、

5 気導マイクロホン1及び生体センサ2からの信号が信号処理回路3へ入力され、また、信号処理回路3からの信号が信号伝送部4を介して外部へ出力されるように電気的に接続されている。

【0018】気導マイクロホン1は、通常の空気振動を検出する、ダイナミック型やコンデンサ型といったマイクロホンであり、音声入力用に用いられている、100～16kHz程度の音域において高い利得が得られるものである。気導マイクロホン1は、無指向性のものとしてもよいが、利用者による使用（装着）位置が固定的であれば、特定の方向（例えば利用者の口の方向）指向性のあるものを用いてもよい。利用者の音声のみを検出し易くなるためである。

【0019】生体センサ2は、話者である利用者装着され、利用者の発声動作を検知する。本実施例では、生体センサ2を、骨伝導マイクロホンにて実現している。骨伝導マイクロホンは、圧電式の振動ピックアップを用いたものであり、利用者の耳の近傍に装着することにより、利用者の発声に基づく頭蓋骨の振動を検出するものである。骨伝導マイクロホンによれば、空気振動によって伝達する外界音は人体によって遮断されるため、装着者の発声した音声のみを検出できる。

【0020】信号処理回路3は、一般的なアナログ信号又はデジタル信号を処理する電気回路であり、遅延回路31、出力選択回路32、オン/オフ判定回路33、閾値調整回路34を備えている。オン/オフ判定回路33は、生体センサ2からの信号が閾値を上回った時点からの一定時間を発声区間として、オンを示す信号を出力選択回路32へ出力する。ここでいう閾値（例えば電圧レベル）の利用者による調整を可能にする回路が、閾値調整回路34である。なお、生体センサ2が複数設けられている場合には、オン/オフ判定回路33は、全部あるいは一部の生体センサ2からの信号が閾値を上回っているか否かで発声区間を検知するようにすればよい。またこの場合、閾値調整回路34は、それぞれの生体センサ2の信号についての閾値を調整可能にすることが考えられる。

【0021】遅延回路31は、気導マイクロホン1から入力される音声信号を一定時間だけ遅延させ、出力選択回路32へ出力する。この遅延時間は、上述したオン/オフ判定回路33により、生体センサ2からの信号が処理される時間である。つまり、この遅延回路31によって、出力選択回路32で処理される気導マイクロホン1からの音声信号と、オン/オフ判定回路33への入力信号である生体センサ2からの信号とを同期させるのである。

【0022】出力選択回路32は、オン/オフ判定回路33からの信号に基づき、遅延回路31から出力された音声信号を信号伝送部4へ出力する。具体的には、オンを示す信号が出力されている間は、遅延回路31から出

力された音声信号をそのまま信号伝送部4へ出力する。一方、オフを示す信号が出力されている間は、遅延回路31から出力された音声信号の利得を下げて信号伝送部4へ出力する。本実施例では、オフを示す信号が出力されている間、利得を下げて音声信号を出力する構成としたが、音声信号の出力レベルを「0」、すなわち全く出力しない構成としてもよい。

【0023】なお、このような信号処理回路3は、必要に応じてアナログ/デジタル変換回路、増幅回路、フィルタ回路などを備える構成としてもよい。例えば生体センサ2とオン/オフ判定回路33との間にフィルタ回路と増幅回路を挿入するという具合である。このようにすれば、発声動作を識別し易い通波帯域の信号成分のみを抽出して増幅することができる。また例えば、生体センサ2からの信号のピーク出力部分の揺らぎの波形を出力に変換する回路などを挿入してもよい。

【0024】信号伝送部4は、信号処理回路3の出力選択回路32からの音声信号を、外部装置へ伝送するための構成である。外部装置としては、例えば音声認識装置、電話機、録音装置などが挙げられる。信号伝送部4は例えば、単純な電気配線、変調器、あるいは、無線送信機などとして実現される。

【0025】このように構成された本信号処理装置の外観の一例を示すが、図2に示す外観図である。つまり、本信号処理装置は、その一例として、一般的なヘッドセット型のマイクロホンと同様に形成することができる。図2では、頭部固定用のアーム50の一方の先端部51に気導マイクロホン1が固定され、本装置を利用者がアーム50を用いて頭部に装着すると、利用者の口元に気導マイクロホン1が配置されるようになっている。

【0026】また、他方の先端部52に生体センサである骨伝導マイクロホンが固定されており、頭部装着時には、利用者の耳の近傍の頭部に密着して配置される。そして、信号処理回路3は耳当て部分53に実装されており、この耳当て部分53の信号処理回路3に信号伝送部4としての電気配線が接続されている。

【0027】なお、ここで示した外観図は一例に過ぎず、上述した気導マイクロホン1、生体センサ2、信号処理回路3、及び信号伝送部4の物理的配置は何等限定されるものではない。次に、本実施例の信号処理装置の動作を説明する。ここでは具体的な音声信号等を示す図3の説明図を参照して説明する。

【0028】まず気導マイクロホン1は、利用者の音声を変換振動から音声信号に変換して出力する。このとき、利用者の音声以外の外界音も音声信号に混入する。この気導マイクロホン1からの出力信号を示すが、図3(a)である。一方、骨伝導マイクロホンとしての生体センサ2は、利用者の骨伝導音を信化して出力する。図3(b)に示す如くである。このとき、空気振動によって伝達する外界音は人体によって遮断され

【0029】そして、このような骨伝導音の信号が閾値を上回っているか否かを、信号処理回路3のオン/オフ判定回路33が繰り返し判断し、閾値を上回っていると判断した時点から一定時間、オンを示す信号を出力する。閾値を上回っていると判断された時点から一定時間内に、再び閾値を上回っていると判断がなされなければ、オフを示す信号を出力する。

【0030】このように判断時点から一定時間をオンとするのは、骨伝導音が疎密波であって必ず零点を過ぎるからであり、単に閾値を上回ったか否かでオン/オフを行うと、オン/オフが繰り返されてしまうためである。したがって、上述の一定時間は、この観点から骨伝導音の信号の最長振動周期(10ms)程度とすればよい。これによって、骨伝導音が継続していれば、この一定時間内に再び閾値を上回ることが輸出され、さらにその検出時点から一定時間、オンを示す信号が出力されることになって、骨伝導音の継続区間、すなわち発声区間に対応させ、オンを示す信号を出力することができる。

【0031】例えば図3(b)に記号 $\alpha$ 、 $\beta$ で示したような信号レベルを閾値として設定すれば、図3(c)に示す区間(斜線を施した区間)において、オン/オフ判定回路33からの信号が出力されることになる。信号処理回路3の出力選択回路32は、図3(c)に示した斜線区間では、遅延回路31から出力される音声信号をそのまま出力し、一方、それ以外の区間では、遅延回路31から出力される音声信号の利得を下げて出力する。図3(d)に示す区間である。

【0032】つまり、図3(a)に示した音声信号に対し、図3(d)に示すような、発声区間以外での利得を下げた音声信号が、出力選択回路32から信号伝送部4を介して出力される。次に、本実施例の信号処理装置の発振する効果を説明する。

【0033】本実施例では、骨伝導マイクホンである生体センサ2からの信号に基づいて、信号処理回路3のオン/オフ判定回路33が、設定された閾値での判断を行い、骨伝導音の継続区間、すなわち利用者の発声区間を検知してオンを示す信号を出力する。そして、出力選択回路32により、オンを示す信号が出力されている区間では、気導マイクホン1からの音声信号がそのまま出力される。これによって、話者が実際に発声している区間、すなわち発声区間の音声信号が出力されることになり、この音声信号を音声認識装置、電話機、あるいは、録音装置への入力信号とすれば、話者の音声の認識が容易になる。特に、音声認識装置において連続した単語を認識する場合に有効である。

【0034】また、本実施例では、オン/オフ判定回路33からオフを示す信号が出力されている間、つまり、発声区間に対応しない区間において、出力選択回路32は、気導マイクホン1からの音声信号の利得を下げて出力する。これによって、区間の境界部分に外界音など

が存在しても、雑音が入る可能性を低減でき、さらに、音声の認識が容易になる。

【0035】なお、本実施例の気導マイクホン1が「気導マイクホン」に相当し、生体センサ2が「生体センサ」に相当する。また、信号処理回路3のオン/オフ判定回路33が「検知手段」に相当し、出力選択回路32及び信号伝送部4が「出力手段」に相当する。

【0036】以上、本発明はこのような実施例に何等限定されるものではなく、本発明の主旨を逸脱しない範囲において種々なる形態で実施し得る。

(イ)上記実施例では、生体センサ2からの骨伝導音に基づいて発声区間を特定するために、オン/オフ判定回路33が、生体センサ2からの信号が閾値を上回った時点から一定時間、オンを示す信号を出力していた。

【0037】これに対して、オン/オフ判定回路33は閾値を上回った時だけにオンを示す信号を出力する構成とし、代わりに、出力選択回路32が、オンを示す信号を検知した時点から一定時間、音声信号をそのまま出力する構成としてもよい。この場合は、オン/オフ判定回路33と共に出力選択回路32が「検知手段」に相当することになる。

【0038】また、生体センサ2からの信号が零点を必ず通るような信号でなければ、生体センサ2からの信号が閾値を上回る時にオンを示す信号をオン/オフ判定回路33からの出力するようにし、出力選択回路32は、オンを示す信号が入力されている間だけ、音声信号をそのまま出力する構成とすることができる。

【0039】(ロ)上記実施例では生体センサ2として骨伝導マイクホンをを用いたが、利用者の発声動作、すなわち、声帯の振動、口の開閉、舌の動き、喉の気道の変化、顎の筋肉の動き等を検知できるセンサであればよい。例えば、圧電振動子を用いた超音波センサを用いて生体センサ2を構成してもよい。この場合、顎の骨や歯筋に密着させて装着することによって、利用者の発声に伴う声帯の振動や舌の動き、喉の気道の変化が検知できる。特に、超音波センサを用いたとき、骨伝導マイクホンでは検知し難い無音声の発声動作を舌の動き、喉の気道の変化から検知できるという点で有利である。

【0040】また例えば、筋電位センサ、加速度センサ、赤外線センサ、発光素子と受光素子とを用いた光式振動ピックアップなどを用いて生体センサ2構成してもよい。これらのセンサを首や耳の下あたりに装着することによって、発声時の声帯の振動に伴う生体の微小振動や、発声時の口、顎の動きを検知できる。特に、筋電位センサや加速度センサは安価であるため、信号処理装置を安価に構成できる。また、赤外線センサ、光式振動ピックアップを用いた場合は、利用者に非接触で検知できるので、密着させて使用されるセンサに比べ、センサの密着による利用者の違和感を小さくできる。

【0041】もちろん、同種類又は異種類のセンサを複

数個用いて、生体センサ2を構成してもよい。

(ハ) なお、骨伝導マイクロホンを用いれば、利用者の発声した音声のみが検出できるため、骨伝導マイクロホンからの音声信号そのものを出力してもよいように思われる。つまり、骨伝導マイクロホンからの音声信号は、もともと発声区間に対応するものであり、外界音が遮断されたものとなっている。

【0042】しかしながら、骨伝導マイクロホンは、気導マイクロホンと比較して、約2 KHz以上の高音域の利得が低く、検出された音声信号には高音域が欠落してしまう。そのため、いわゆる「こもった」音声となり、不明瞭で認識し難い。したがって、本発明では、発声区間を特定するための生体センサの一つとして用い、実際に出力する音声信号は、気導マイクロホン1からの音声信号として、その明瞭性を確保するのである。

【図面の簡単な説明】

【図1】実施例の信号処理装置の概略構成を示すブロック図である。

【図2】信号処理装置の外観の一例を示す説明図である。

【図3】音声信号の処理を示すための説明図である。

【符号の説明】

1…気導マイクロホン

2…生体センサ

3…信号処理回路

4…信号伝送部

31…遅延回路

32…出力選択回路

33…オン／オフ判定回路

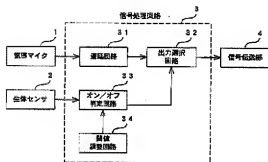
34…閾値調整回路

50…アーム

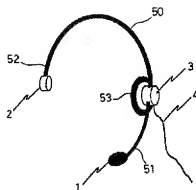
51、52…先端部

53…耳当て部分

【図1】



【図2】



【図3】

